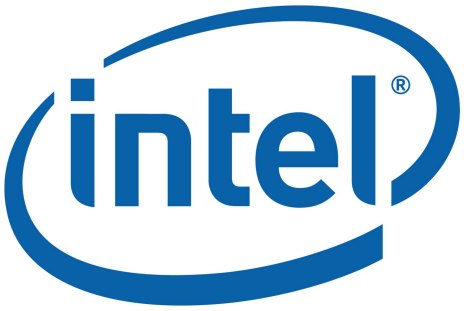


Intel® Ethernet Switch FM5000/ FM6000

1 Gb/2.5 Gb/10 Gb/40 Gb Ethernet (GbE) L2/L3/L4 Chip

Specification Update

Networking Division (ND)

April 2014

Revision 1.6

LEGAL

By using this document, in addition to any agreements you have with Intel, you accept the terms set forth below.

You may not use or facilitate the use of this document in connection with any infringement or other legal analysis concerning Intel products described herein. You agree to grant Intel a non-exclusive, royalty-free license to any patent claim thereafter drafted which includes subject matter disclosed herein.

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

A “Mission Critical Application” is any application in which failure of the Intel Product could result, directly or indirectly, in personal injury or death. SHOULD YOU PURCHASE OR USE INTEL'S PRODUCTS FOR ANY SUCH MISSION CRITICAL APPLICATION, YOU SHALL INDEMNIFY AND HOLD INTEL AND ITS SUBSIDIARIES, SUBCONTRACTORS AND AFFILIATES, AND THE DIRECTORS, OFFICERS, AND EMPLOYEES OF EACH, HARMLESS AGAINST ALL CLAIMS COSTS, DAMAGES, AND EXPENSES AND REASONABLE ATTORNEYS' FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PRODUCT LIABILITY, PERSONAL INJURY, OR DEATH ARISING IN ANY WAY OUT OF SUCH MISSION CRITICAL APPLICATION, WHETHER OR NOT INTEL OR ITS SUBCONTRACTOR WAS NEGLIGENT IN THE DESIGN, MANUFACTURE, OR WARNING OF THE INTEL PRODUCT OR ANY OF ITS PARTS.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked “reserved” or “undefined”. Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or go to: http://www.intel.com/design/literature.htm.

Intel and Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

* Other names and brands may be claimed as the property of others.

Revision History

Revision	Date	Comments
1.6	April 23, 2014	**Documentation Updates added:** • 1. I2C Clock Divider (Added) • 2. EPL Standard Modes (Added)
1.5	November 14, 2013	**Specification Clarifications added or updated:** • 1. Thermal Stress Testing (Added) **Other updates:** • Added component ordering information. • Reformatted document to current Intel corporate standards.
1.2	July 2013	Initial release. (Intel confidential)

NOTE: ***This page intentionally left blank.***

1.0 Introduction

This document applies to the Intel® Ethernet Switch FM5000/FM6000.

This document is an update to a published specification, the *Intel® Ethernet Switch FM5000/FM6000 Datasheet*. It is intended for use by system manufacturers and software developers. All product documents are subject to frequent revision and new order numbers might apply. New documents might be added. Be sure you have the latest information before finalizing your design.

1.1 Product Code and Device Identification

Product Codes:

The following tables and drawings describe the various identifying markings on each device package:

Table 1-1 Markings

Device	Stepping	Top Marking	Q-Specification	Description
Intel® Ethernet Switch FM6724	B2	EZFM6724A	SLKAA	24 10 Gb/s ports or 6 40 Gb/s ports, fully-integrated, single-chip wire-speed, layer-2/3/4 Ethernet switch with SDN enhancements.

Table 1-2 MM Numbers

Product	Tray MM#
Intel® Ethernet Switch FM6724	930424

1.2 Marking Diagrams

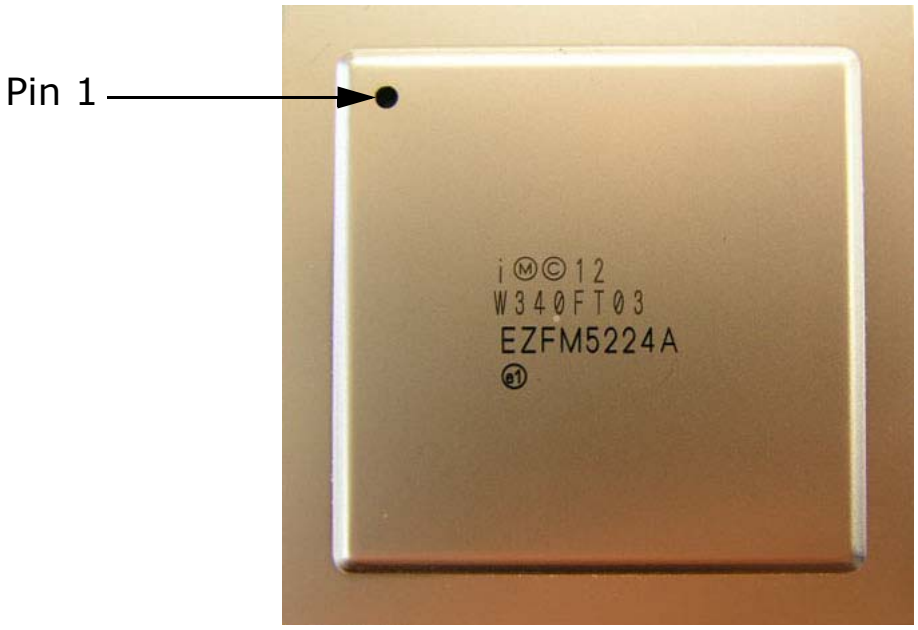


Figure 1-1. FM5000 Example With Identifying Marks

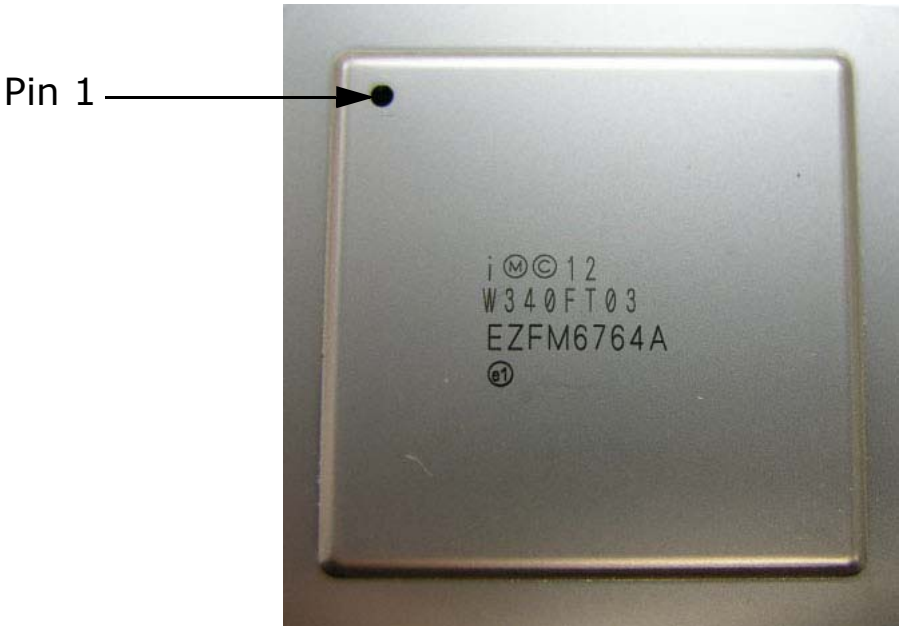
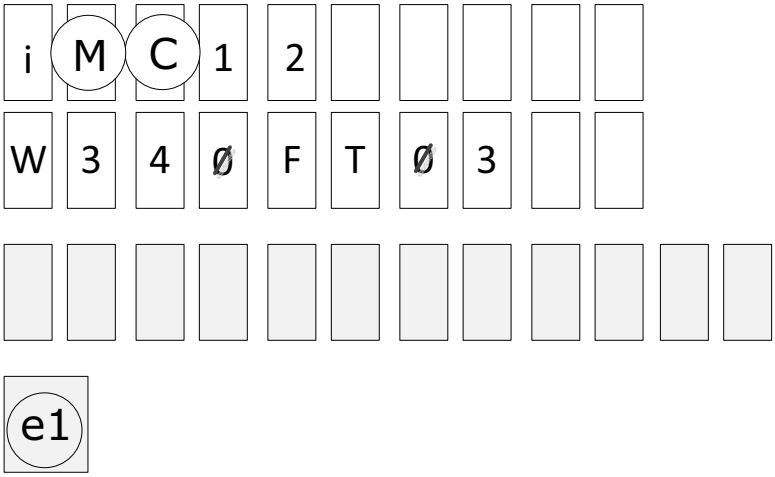


Figure 1-2. FM6000 Example With Identifying Marks



- LINE1: i, maskwork, and copyright
- LINE2: FPO number
- LINE3: Product name followed by test designator (C, E, or I). Left blank by assembly sub-contractor.
- LINE4: Pb free symbol

1.3 Nomenclature Used in This Document

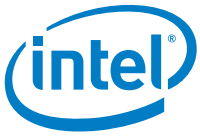
This document uses specific terms, codes, and abbreviations to describe changes, errata, sightings and/or clarifications that apply to silicon/steppings. See Table 1-1 for a description.

Table 1-1 Nomenclature

Name	Description
Specification Clarifications	Greater detail or further highlights concerning a specification's impact to a complex design situation. These clarifications will be incorporated in the next release of the specifications.
Specification Changes	Modifications to the current published specifications. These changes will be incorporated in the next release of the specifications.
Errata	Design defects or errors. Errata may cause device behavior to deviate from published specifications. Hardware and software designed to be used with any given stepping must assume that all errata documented for that stepping are present on all devices.
Software Clarifications	Applies to Intel drivers, EEPROM loads.
Documentation Changes	Typos, errors, or omissions from the current published specifications. These changes will be incorporated in the next release of the specifications.
A1, B1, etc.	Stepping to which the status applies.
Doc	Document change or update that will be implemented.

Table 1-1 Nomenclature (Continued)

Fix	This erratum is intended to be fixed in a future stepping of the component.
Fixed	This erratum has been previously fixed.
NoFix	There are no plans to fix this erratum.
Eval	Plans to fix this erratum are under evaluation.
Red Change Bar/or Bold	This Item is either new or modified from the previous version of the document.



2.0 Hardware Clarifications, Changes, Updates and Errata

See Section 1.3 for an explanation of terms, codes, and abbreviations.

Table 2-1 Summary of Specification Clarifications

Specification Clarification	Status
1. Thermal Stress Testing	N/A

Table 2-2 Summary of Specification Changes

Specification Change	Status
None.	N/A

Table 2-3 Summary of Documentation Updates

Documentation Update	Status
1. I2C Clock Divider	N/A
2. EPL Standard Modes	

Table 2-4 Summary of Errata Including Steppings

Erratum	Status
1. Transmit Skew Higher Than 10GBASE-CX4 Requirement	B2; NoFix
2. L2AR SRAM Errors Might Corrupt Policers and Counters	B2; NoFix
3. PCIe Minimum Frame Size Padding Deadlock	B2; NoFix
4. PCIe Interface Impedance Not Compliant at Power Up	B2; NoFix
5. Small Sized Frames That Are Sent to the CPU Might Not Timeout	B2; NoFix

Table 2-4 Summary of Errata Including Steppings (Continued)

Erratum	Status
6. Deficit Round Robin Scheduling Fails in Presence of L3 Multicast Replication	B2; NoFix
7. Independent Lane Reversal Not Supported	B2; NoFix
8. Device Lock-up For More Than 64 Ports	B2; NoFix
9. MDIO Pin Is Not Open Drain	B2; NoFix
10. Power Supply Scaling Requirement	B2; NoFix
11. Inconsistent Statistics	B2; NoFix
12. PCIe Single Lane DMA Issue	B2; NoFix

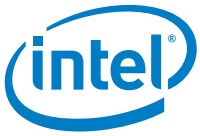
2.1 Specification Clarifications

1. Thermal Stress Testing

For thermal stress testing, Intel recommends using the conditions described in section 9.2.3 of the *Intel® Ethernet Switch FM5000/FM6000 Datasheet* for TDP (maximum sustained power). These conditions include a combined interleaved traffic profile that consists of 5% of 64-byte + 75% of 256-byte packets while maintaining a Tcase of 75 °C for a typical TDP stress condition and a Tcase of 85 °C for margined TDP stress conditions. Extended testing outside of these conditions or for long periods of time (as long as 2 days) might reduce the total lifetime of the part such that full operating lifetime of 7 years cannot be guaranteed.

2.2 Specification Changes

None.



2.3 Documentation Updates

Note: The following updates will appear in the next revision of the *Intel® Ethernet Switch FM5000/FM6000 Datasheet*.

1. I2C Clock Divider

The following update applies to the I2C Clock Divider within the I2C_CFG register (Section 10.35.2.23).

Field Name	Bit(s)	Type	Default	Description
Divider	19:8	RW	0xC	I2C Clock Divider. The I2C clock is equal to PCIE_REFCLK/19/DIVIDER.

2. EPL Standard Modes

Table 6-6, "EPL Standard Modes" contains the following updates:

Mode	#SerDes	Standard	Encoding
SFP+ Direct Attach Copper[1]	1	SFF 8431	64b/66b
SFP+ SFI Interface (10GBASE-SR)[2]	1	SFF 8431	64b/66b

1. For SFP+ Direct Attach applications, Intel recommends using 24 gauge direct attach cables that are 5 meters or less.
2. Deterministic jitter generation is 150 mUI maximum verses SFI specification of 100 mUI maximum.

2.4 Errata

See Section 1.3 for an explanation of terms, codes, and abbreviations.

1. Transmit Skew Higher Than 10GBASE-CX4 Requirement

Severity:

Low

Problem:

The worst case transmit skew for PCS10GBASE-CX4 is 20-UI; however, the specification is set as one Unit Interval (UI), therefore the transmit module is not 100% compliant with the skew budget specification presented in the IEEE 802.3—Table 48-5 Skew Budget table.

Implication:

The transmitter creates a skew higher than what is allowed in the 10GBASE-CX4 specification. This might prevent a link from being established if the total skew is greater than what the receiver can tolerate. Such a situation can arise when a FM5000/FM6000 is connected to another device using an interconnection system that introduces a large skew between lanes. In such a scenario, the added skew might cause the receiving devices' realignment buffer to overflow.

Workaround:

This issue is only expected to manifest itself when the device is used with 10GBASE-CX4. 40G/100G Base-R are more tolerant protocols that enable greater skew flexibility. For example, consider a case in which the FM5000/FM6000 is connected to another device via XAUI/CX4. If the skew between the lanes introduced by the interconnection between those devices is virtually zero (like a CX4 cable or PCB traces of equal length), then the 20 UI skew that is introduced by the transmitter does not cause any problems as the skew is well below the 40 UI that the receiver is expected to support.

Status: B2; NoFix

2. L2AR SRAM Errors Might Corrupt Policers and Counters

Severity:

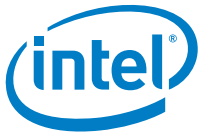
Medium

Problem:

SRAM errors in the L2AR block might cause statistical index errors as well as policer counter errors.

Implication:

If a frame hits an SRAM entry in the L2AR block that has a soft error, the error is reported to management, however, the error is ignored by the forwarding hardware down the pipeline potentially resulting in errors in the statistical counters and policer counters.



Workaround:

A software/API workaround is under investigation.

Status: B2; NoFix

3. PCIe Minimum Frame Size Padding Deadlock

Severity:

Low

Problem:

The PCIe interface does not correctly mark the sub-segment marker when the frame data terminates at a sub-segment boundary and the frame must be padded to the minimum frame size. The effect of missing the subsegment marker is that the ingress crossbar deadlocks and stalls the PCIe port channel.

Implication:

The PCIe interface might send unpadded frames to the ingress crossbar resulting in the crossbar/PCIe interface deadlocking.

Workaround:

The software workaround is to set PCI_TX_FRAME_LEN.MinLen to a value less that the smallest expected frame size (such as 16) and configure the switch to pad the frame as it egresses the switch.

Status: B2; NoFix

4. PCIe Interface Impedance Not Compliant at Power Up

Severity:

Low

Problem:

The PCIe specification defines that the impedance to ground (Zrx-dc) must be approximately 40-to-60 Ω at power up. The FM5000/FM6000 does not comply with that statement.

Implication:

The PCIe specification defines that the impedance to ground (Zrx-dc) must be approximately 40-to-60 Ω at power up. The specification states that the Rx DC single-ended impedance must be present when the receiver terminations are first enabled to ensure that receiver detect occurs properly. The FM5000/FM6000 is not compliant with that statement as it only presents an impedance of 75 Ω.

Workaround:

To ensure the correct impedance, the switch has to be taken out of reset and a minimum amount of configuration is required. In particular, the FM5000/FM6000 should be taken out of reset shortly after power up and a SERIAL SPI boot should be used to minimally configure the switch to get PCIe to boot and provide the correct impedance to establish a link. The configuration sequence performed is a minimal EEPROM image, which can be obtained from Intel.

Status: B2; NoFix

5. Small Sized Frames That Are Sent to the CPU Might Not Timeout

Severity:

Low

Problem:

Frames that are sent to the CPU through the internal MSB block causes the MSB to notify the CPU of the presence of a frame. If the CPU does not retrieve the frame in a certain amount of time since this notification and the frame is smaller than 253 bytes, the timeout mechanism might not timeout the frame resulting in the MSB buffer to remain full.

Implication:

Once a frame destined to the CPU has not timed-out, subsequent frames do not reach the CPU as the MSB remains full.

Workaround:

The CPU should periodically read and try to empty the MSB.

Status: B2; NoFix

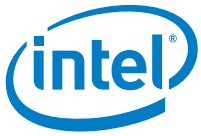
6. Deficit Round Robin Scheduling Fails in Presence of L3 Multicast Replication

Severity:

Medium

Problem:

Deficit round robin scheduling no longer works when the FM5000/FM6000 starts performing layer 3 multicast replication. The scheduling algorithm no longer performs as expected and does not provide the shaping function that is programmed. Furthermore, even after layer 3 multicast replication completes, the FM5000/FM6000 cannot recover from this corrupt state.



Implication:

Deficit round robin scheduling used for shaping purposes might not produce the desired shaping functionality in the presence of layer 3 multicast replication.

Workaround:

None

Status: B2; NoFix

7. Independent Lane Reversal Not Supported

Severity:

Medium

Problem:

Each 2.5 Gb/s port lane consists of an Rx (receive) and Tx (transmit) paths. Lanes can be reversed when each lane's Rx and Tx path is kept together, however, lane reversal is not supported when Rx and Tx paths are treated separately.

Consider the following example:

Lane0	Lane1	Lane2	Lane3
Tx0 Rx0	Tx1 Rx1	Tx2 Rx2	Tx3 Rx3

The following reversal is not supported:

Lane0	Lane1	Lane2	Lane3
Tx0 Rx3	Tx1 Rx2	Tx2 Rx1	Tx3 Rx0

Whereas the following reversal is supported:

Tx3 Rx3	Tx2 Rx2	Tx1 Rx1	Tx0 Rx0

Implication:

Failure to do follow the correct lane reversal method might result in incorrect operation.

Workaround:

None

Status: B2; NoFix

8. Device Lock-up For More Than 64 Ports

Severity:

High

Problem:

The FM5000/FM6000 might lock up when subjected to maximum traffic loads on more than 64 10 GbE ports (or any combination of 10 GbE and 40 GbE ports equivalent to 65 or more 10 GbE ports).

Implication:

Above a sustained load of 640 Gb/s with small packet sizes, the FM5000/FM6000 might lock up and require a hard reset before it can process further management or packet activity. The severity of this problem depends on the operational voltage and temperature.

Workaround:

A software workaround that converts the lock-up failure mode to a packet dropping failure mode is available. This patch is available starting in SDK version 3.3.0. With this workaround in place, packets might be dropped only under periods of sustained high load with small packets.

Please contact your local Intel representative for more information.

Status: B2; NoFix

9. MDIO Pin Is Not Open Drain

Severity:

Low

Problem:

The MDIO output pin is specified in the *Intel® Ethernet Switch FM5000/FM6000 Datasheet* as open drain but it actually has an active pull-up resistor to the VDD25 supply.

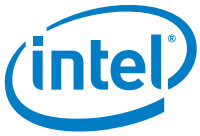
Implication:

The MDIO output pin does not behave as an actual open-drain output during some of its operation. During parts of the transaction, the pin actively pulls high to VDD25, but then behaves as open-drain during the remainder of the transaction. Because of this, the output still needs an external pull-up resistor to VDD25.

Workaround:

If the MDIO usage is point-to-point there should be no problem. If the signal is expected to be shared with other outputs, customers should use a level translator/isolator device, like the Pericom* PI4ULS3V08 between the FM5000/FM6000 and the shared signal line.

Status: B2; NoFix



10. Power Supply Scaling Requirement

Severity:

Medium

Problem:

VDD and VDDS power supply scaling is required to meet the FM5000/FM6000 performance requirements.

Implication:

When FM5000/FM6000 Ethernet switches are tested, the VDD and VDDS power supply values are programmed in the fuse box. Software must read the fuse box and then configure the programmable power supplies accordingly.

Workaround:

An API function call is available, in a patch to release 3.3.1, to read these numbers from the fuse box and convert them into power supply voltage values that can be used by the customer's software to configure the power supplies.

Status: B2; NoFix

11. Inconsistent Statistics

Severity:

Low

Problem:

Statistics for truncated frames might be incorrectly reported.

Implication:

Truncation in the EPL to TxMinColumns was disabled in B0, but frames still get truncated to a maximum of 160 bytes by the data path. However, statistics still uses MOD_MIN_LENGTH for both padding and truncation, so the statistics for one or the other is wrong (unless they are padded to 160 bytes). For example, a 100-byte frame that is marked for truncation egresses as a 100-byte frame, but is counted as a 64-byte frame if MOD_MIN_LENGTH=64.

Workaround:

The workaround consists in setting the FM_PORT_TX_PAD_SIZE to a value of 160 bytes. By doing this, all truncated frames (either <160 bytes or ≥160 bytes) always have a length of 160 bytes and size counters match.

Status: B2; NoFix

12. PCIe Single Lane DMA Issue

Severity:

Low

Problem:

When operating the PCIe management interface in single-lane mode, the interface can hang during DMA transfers. If the following conditions occur, the PCIe interface hangs:

- Only affects Tx DMA (not Rx)
- Current Buffer Descriptor (BD) did not contain a STOP flag
- There are no more valid BDs internally, so a fetch must occur next.

Implication:

The PCIe management interface might become inoperative requiring the chip to be reset. When operating in PCIe 4-lane mode, this issue cannot occur.

Workaround:

A software workaround consists of forcing the DMA engine to transition to the STOP state after executing a Tx DMA transaction, in order to prevent fetching a new BD. At the device driver level, that can be achieved by setting the STOP bit in the BD status field, when the BD is marked as READY. Refer to the *Intel® FM5000/FM6000 Series Ethernet Switch and Router Device Erratum #12 Workaround* document for more details.

Status: B2; NoFix